RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/825,692_

Source: _1FWO_

Date Processed by STIC: _11/8/04_

ENTERED

RAW SEQUENCE LISTING   DATE: 11/08/2004
PATENT APPLICATION: US/10/825,692   TIME: 16:09:13

      Input Set : A:\03740007aa.txt
      Output Set: N:\CRF4\11082004\J825692.raw

 3 <110> APPLICANT: Hotez, Peter
 4   Ashcom, James
 5   Bdamchian, Mahnaz
 6   Zhan, Bin
 7   Wang, Yan
 8   Hawdon, John
 9   Loukas, Alexander
10   Williamson, Angela
11   Jones, Brian
12   . Bethony, Jeffrey
13   Goud, Gaddam
14   Botazzi, Maria E.
15   Mendez, Susana
17 <120> TITLE OF INVENTION: Hookworm Vaccine
19 <130> FILE REFERENCE: 03740007aa
C--> 21 <140> **CURRENT APPLICATION NUMBER: US/10/825,692**
C--> 21 <141> **CURRENT FILING DATE: 2004-04-16**
21 <150> PRIOR APPLICATION NUMBER: US 60/329,533
22 <151> PRIOR FILING DATE: 2001-10-17
24 <150> PRIOR APPLICATION NUMBER: US 60/332,007
25 <151> PRIOR FILING DATE: 2001-11-23
27 <150> PRIOR APPLICATION NUMBER: US 60/375,404
28 <151> PRIOR FILING DATE: 2002-04-26
30 <150> PRIOR APPLICATION NUMBER: PCT US02/33106
31 <151> PRIOR FILING DATE: 2002-10-17
33 <160> NUMBER OF SEQ ID NOS: 114
35 <170> SOFTWARE: PatentIn version 3.2
37 <210> SEQ ID NO: 1
38 <211> LENGTH: 1451
39 <212> TYPE: DNA
40 <213> ORGANISM: Necator americanus
42 <400> SEQUENCE: 1
43 atgttttctc ctgtagtcgt cagtgtggta ttcacaatcg ccttctgcaa tgcgtctcca   60
45 gcaagagaca gcttcggctg ctctaacagt gggataactg acagcgaccg gcaagcgttc   120
47 ctcgacttcc acaacaatgc tcgtcgacgg gttgcgaaag gccttgagga tagcaactcc   180
49 ggcaaactga atccagcgaa gaacatgtac aagctgtcat gggactgtgc aatggaacag   240
51 cagcttcagg atgccatcca gtcatgccca agcggctttg ctgggattca aggtgttgcg   300
53 cagaatacaa tgagctggtc aagctctggt ggataccccg atccatcggt aaagatagaa   360
55 ccaacgctct ccggctggtg gagtggtgcg aaaaagaacg gcgtaggccc ggacaacaaa   420
57 tacaccggtg gtggtctctt cgccttctct aacatggtat actccgaaac gacgaaactt   480
59 ggctgcgctt acaaggtttg·cggcactaaa ctggcggttt catgcatcta taatggagtc   540
61 gggtacatca caaatcaacc tatgtgggag acaggtcagg cttgccagac aggagcagac   600
63 tgctccactt acaagaactc aggctgcgag gacggccttt gcacgaaggg accagatgta   660

```
65 ccagaaacaa accagcagtg cccctcaaac accggaatga ctgattcagt cagagatact    720
67 ttcctatcgg tgcacaatga gttcagatcg agtgttgccc gaggtctgga acccgacgct    780
69 ctgggcggaa atgcaccaaa agcagctaaa atgctcaaga tggtgtatga ctgtgaagtg    840
71 gaagcatcgg ccatcagaca tggaaataaa tgcgtctatc aacattctca tggtgaagac    900
73 agacctggac taggagaaaa catctacaaa actagtgtac tcaaattcga caagaacaaa    960
75 gcagccaagc aggcttcaca actctggtgg aatgagttaa aagagtacgg cgtcggccca   1020
77 tccaacgtcc ttaccactgc gttatggaat agacccaaca tgcagattgg tcactacacc   1080
79 cagatggcat gggacaccac ctacaaactt ggatgtgcag ttgttttctg caatgatttc   1140
81 acattcggcg tttgtcagta tgggccagga ggcaattaca tgggtcatgt catctacact   1200
83 atgggccagc cgtgctctca gtgttcgcct ggtgctactt gcagcgtgac cgaaggcttg   1260
85 tgcagcgctc cttaatcagt caacaataaa tatcttacag tgatgttgtt gcttacaaat   1320
87 tgcttctttt ccaatagaaa taccaatgtc aacatcacga gtttctttaa attcatcact   1380
89 tccactacta ggggtgattt gaataaaatt tcatttcata aagcaattac atccgcaaaa   1440
91 aaaaaaaaaa a                                                          1451
94 <210> SEQ ID NO: 2
95 <211> LENGTH: 424
96 <212> TYPE: PRT
97 <213> ORGANISM: Necator americanus
99 <400> SEQUENCE: 2
101 Met Phe Ser Pro Val Val Val Ser Val Val Phe Thr Ile Ala Phe Cys
102 1               5                   10                  15
105 Asn Ala Ser Pro Ala Arg Asp Ser Phe Gly Cys Ser Asn Ser Gly Ile
106             20                  25                  30
109 Thr Asp Ser Asp Arg Gln Ala Phe Leu Asp Phe His Asn Asn Ala Arg
110         35                  40                  45
113 Arg Arg Val Ala Lys Gly Leu Glu Asp Ser Asn Ser Gly Lys Leu Asn
114     50                  55                  60
117 Pro Ala Lys Asn Met Tyr Lys Leu Ser Trp Asp Cys Ala Met Glu Gln
118 65                  70                  75                  80
121 Gln Leu Gln Asp Ala Ile Gln Ser Cys Pro Ser Gly Phe Ala Gly Ile
122             85                  90                  95
125 Gln Gly Val Ala Gln Asn Thr Met Ser Trp Ser Ser Ser Gly Gly Tyr
126             100                 105                 110
129 Pro Asp Pro Ser Val Lys Ile Glu Pro Thr Leu Ser Gly Trp Trp Ser
130         115                 120                 125
133 Gly Ala Lys Lys Asn Gly Val Gly Pro Asp Asn Lys Tyr Thr Gly Gly
134     130                 135                 140
137 Gly Leu Phe Ala Phe Ser Asn Met Val Tyr Ser Glu Thr Thr Lys Leu
138 145                 150                 155                 160
141 Gly Cys Ala Tyr Lys Val Cys Gly Thr Lys Leu Ala Val Ser Cys Ile
142             165                 170                 175
145 Tyr Asn Gly Val Gly Tyr Ile Thr Asn Gln Pro Met Trp Glu Thr Gly
146             180                 185                 190
149 Gln Ala Cys Gln Thr Gly Ala Asp Cys Ser Thr Tyr Lys Asn Ser Gly
150             195                 200                 205
153 Cys Glu Asp Gly Leu Cys Thr Lys Gly Pro Asp Val Pro Glu Thr Asn
154     210                 215                 220
157 Gln Gln Cys Pro Ser Asn Thr Gly Met Thr Asp Ser Val Arg Asp Thr
158 225                 230                 235                 240
```

RAW SEQUENCE LISTING   DATE: 11/08/2004
PATENT APPLICATION: US/10/825,692   TIME: 16:09:13

Input Set : A:\03740007aa.txt
Output Set: N:\CRF4\11082004\J825692.raw

```
161 Phe Leu Ser Val His Asn Glu Phe Arg Ser Ser Val Ala Arg Gly Leu
162             245                 250                 255
165 Glu Pro Asp Ala Leu Gly Gly Asn Ala Pro Lys Ala Ala Lys Met Leu
166             260                 265                 270
169 Lys Met Val Tyr Asp Cys Glu Val Glu Ala Ser Ala Ile Arg His Gly
170             275                 280                 285
173 Asn Lys Cys Val Tyr Gln His Ser His Gly Glu Asp Arg Pro Gly Leu
174         290                 295                 300
177 Gly Glu Asn Ile Tyr Lys Thr Ser Val Leu Lys Phe Asp Lys Asn Lys
178 305                 310                 315                 320
181 Ala Ala Lys Gln Ala Ser Gln Leu Trp Trp Asn Glu Leu Lys Glu Tyr
182             325                 330                 335
185 Gly Val Gly Pro Ser Asn Val Leu Thr Thr Ala Leu Trp Asn Arg Pro
186             340                 345                 350
189 Asn Met Gln Ile Gly His Tyr Thr Gln Met Ala Trp Asp Thr Thr Tyr
190         355                 360                 365
193 Lys Leu Gly Cys Ala Val Val Phe Cys Asn Asp Phe Thr Phe Gly Val
194     370                 375                 380
197 Cys Gln Tyr Gly Pro Gly Gly Asn Tyr Met Gly His Val Ile Tyr Thr
198 385                 390                 395                 400
201 Met Gly Gln Pro Cys Ser Gln Cys Ser Pro Gly Ala Thr Cys Ser Val
202             405                 410                 415
205 Thr Glu Gly Leu Cys Ser Ala Pro
206             420
209 <210> SEQ ID NO: 3
210 <211> LENGTH: 1893
211 <212> TYPE: DNA
212 <213> ORGANISM: Necator americanus
214 <400> SEQUENCE: 3
215 ggtactgcag ggtttaatta cccaagtttg agacccaacg ccatgatttg gcgaacgtgg    60
217 caagttctcg tggttctgta tgcggcgctg tccattacag ttgtgaacgc ctataaacac   120
219 attagctccg atcacgttgt aaatacaaca ctgggtcaga ttcgaggagt accacagaat   180
221 ttcgaaggca aaaagttac cgcttttctt ggtgtgccat atggtcaacc accgactggg   240
223 gaactacgat tcagcaatcc gaaaatggtg cagcgttggg aaggtataaa gaatgctaca   300
225 acaccggctc agccatgctt ccacttccct gacagtaaat ttaagggatt tcgtgggtca   360
227 gagatgtgga atccgaaagg aaatatgacc gaggattgct tgaatatgaa tatctgggtc   420
229 ccacacgatg ctgatggttc cgtgattgta tggattttcg gaggcggctt cttcaccggt   480
231 tcaccatctt tagatgttta caacggtact gctctagcag ccaagaaacg taccattgtt   540
233 gtgaacataa actatcgatt gggtcccttc ggtttccttt atctcggtga tgattctcgt   600
235 gcacaaggga atatgggact gcaagatcaa caagttgcat tgcgatgggt gcataaacat   660
237 ataagctcct ttggtggaga tccgagaaaa gtcactcttt cggcgaagc atcaggcgct   720
239 gcttcagcaa ccgctcatct agcagcaccg ggaagctatg agttttcga taagataatt   780
241 ggcaacggtg gcacaatcat gaatagttgg gccagtcgaa caaatacatc gatgcttgag   840
243 ctgtcaatga aacttgctga acggttgaac tgtaccaaga aagaaaaga cccgaatact   900
245 gtacatcgct gtttggttaa acatccagca catgtggttc taaaagaggc cgctgttgtg   960
247 tcgtatcaaa ttggtctcgt gctgacgttt gccttcatac ccattacctc tgataagaac  1020
249 ttcttccagg gaaatgtctt tgatcgtcta cgagataaag acattaagaa gaatgtatcc  1080
251 attgtgcttg gtactgtaaa agacgaagca accttctttt taccctacta ctttggtcac  1140
253 aacggtttct cttttcaataa ctcattctta gcagatgggg aagaaaacag agcactcata  1200
```

```
255 aatatatcac agtataatta tgcgatgaat gcaactgcgc catcacttga aagctcactg   1260
257 gaaccacttt tagaagctta taagaacgtt tcgacgcgaa aagaagaagg tgaaagatta   1320
259 cgcgatggtg ttggtcgatt catgggcgac tacttctata cctgcagcgt cattgatttc   1380
261 gctaatatcg tctcagacat tattaatgga tctttgtata tgtattactt tactaagagg   1440
263 tcagtggcaa atccttggcc agagtggatg ggtgtaatgc atggttatga aatagaatac   1500
265 gaatttggac agcctttcct aaattcatca ctgtacaagg aaaagcttga aaacgaaaag   1560
267 atcttctcga aaaatatcat gagcttttgg aaagatttca tcaagactgg tgtccctgtc   1620
269 gattttggc cgaaatacga tcgaaaggag cggaaagcgc tcgtacttgg cgaggaaagc    1680
271 gtgaacaatt cttaccctaa tatgactaat gttcatggac cgtactgtga actgatcgaa   1740
273 gaagcaaagg cgtctacaaa taatggactc accttgaaga aatacattga aggggagata   1800
275 aaaaataacg aaacgaacgt attttgatag aatgattttg cacagaatga agaattgaat   1860
277 atcaaaaaaa aaaaaaaaa aaaaaaaaa aaa                                   1893
280 <210> SEQ ID NO: 4
281 <211> LENGTH: 594
282 <212> TYPE: PRT
283 <213> ORGANISM: Necator americanus
285 <400> SEQUENCE: 4
287 Met Ile Trp Arg Thr Trp Gln Val Leu Val Val Leu Tyr Ala Ala Leu
288 1               5                   10                  15
291 Ser Ile Thr Val Val Asn Ala Tyr Lys His Ile Ser Ser Asp His Val
292              20                  25                  30
295 Val Asn Thr Thr Leu Gly Gln Ile Arg Gly Val Pro Gln Asn Phe Glu
296              35                  40                  45
299 Gly Lys Lys Val Thr Ala Phe Leu Gly Val Pro Tyr Gly Gln Pro Pro
300          50                  55                  60
303 Thr Gly Glu Leu Arg Phe Ser Asn Pro Lys Met Val Gln Arg Trp Glu
304 65                  70                  75                  80
307 Gly Ile Lys Asn Ala Thr Thr Pro Ala Gln Pro Cys Phe His Phe Pro
308              85                  90                  95
311 Asp Ser Lys Phe Lys Gly Phe Arg Gly Ser Glu Met Trp Asn Pro Lys
312              100                 105                 110
315 Gly Asn Met Thr Glu Asp Cys Leu Asn Met Asn Ile Trp Val Pro His
316          115                 120                 125
319 Asp Ala Asp Gly Ser Val Ile Val Trp Ile Phe Gly Gly Gly Phe Phe
320      130                 135                 140
323 Thr Gly Ser Pro Ser Leu Asp Val Tyr Asn Gly Thr Ala Leu Ala Ala
324 145                 150                 155                 160
327 Lys Lys Arg Thr Ile Val Val Asn Ile Asn Tyr Arg Leu Gly Pro Phe
328              165                 170                 175
331 Gly Phe Leu Tyr Leu Gly Asp Asp Ser Arg Ala Gln Gly Asn Met Gly
332              180                 185                 190
335 Leu Gln Asp Gln Gln Val Ala Leu Arg Trp Val His Lys His Ile Ser
336          195                 200                 205
339 Ser Phe Gly Gly Asp Pro Arg Lys Val Thr Leu Phe Gly Glu Ala Ser
340      210                 215                 220
343 Gly Ala Ala Ser Ala Thr Ala His Leu Ala Ala Pro Gly Ser Tyr Glu
344 225                 230                 235                 240
347 Phe Phe Asp Lys Ile Ile Gly Asn Gly Gly Thr Ile Met Asn Ser Trp
348              245                 250                 255
```

```
351 Ala Ser Arg Thr Asn Thr Ser Met Leu Glu Leu Ser Met Lys Leu Ala
352         260             265             270
355 Glu Arg Leu Asn Cys Thr Lys Lys Arg Lys Asp Pro Asn Thr Val His
356         275             280             285
359 Arg Cys Leu Val Lys His Pro Ala His Val Val Leu Lys Glu Ala Ala
360     290             295             300
363 Val Val Ser Tyr Gln Ile Gly Leu Val Leu Thr Phe Ala Phe Ile Pro
364 305             310             315             320
367 Ile Thr Ser Asp Lys Asn Phe Phe Gln Gly Asn Val Phe Asp Arg Leu
368             325             330             335
371 Arg Asp Lys Asp Ile Lys Lys Asn Val Ser Ile Val Leu Gly Thr Val
372             340             345             350
375 Lys Asp Glu Ala Thr Phe Phe Leu Pro Tyr Tyr Phe Gly His Asn Gly
376         355             360             365
379 Phe Ser Phe Asn Asn Ser Phe Leu Ala Asp Gly Glu Glu Asn Arg Ala
380     370             375             380
383 Leu Ile Asn Ile Ser Gln Tyr Asn Tyr Ala Met Asn Ala Thr Ala Pro
384 385             390             395             400
387 Ser Leu Glu Ser Ser Leu Glu Pro Leu Leu Glu Ala Tyr Lys Asn Val
388             405             410             415
391 Ser Thr Arg Lys Glu Glu Gly Glu Arg Leu Arg Asp Gly Val Gly Arg
392         420             425             430
395 Phe Met Gly Asp Tyr Phe Tyr Thr Cys Ser Val Ile Asp Phe Ala Asn
396         435             440             445
399 Ile Val Ser Asp Ile Ile Asn Gly Ser Leu Tyr Met Tyr Tyr Phe Thr
400     450             455             460
403 Lys Arg Ser Val Ala Asn Pro Trp Pro Glu Trp Met Gly Val Met His
404 465             470             475             480
407 Gly Tyr Glu Ile Glu Tyr Glu Phe Gly Gln Pro Phe Leu Asn Ser Ser
408             485             490             495
411 Leu Tyr Lys Glu Lys Leu Glu Asn Glu Lys Ile Phe Ser Lys Asn Ile
412         500             505             510
415 Met Ser Phe Trp Lys Asp Phe Ile Lys Thr Gly Val Pro Val Asp Phe
416         515             520             525
419 Trp Pro Lys Tyr Asp Arg Lys Glu Arg Lys Ala Leu Val Leu Gly Glu
420     530             535             540
423 Glu Ser Val Asn Asn Ser Tyr Pro Asn Met Thr Asn Val His Gly Pro
424 545             550             555             560
427 Tyr Cys Glu Leu Ile Glu Glu Ala Lys Ala Ser Thr Asn Asn Gly Leu
428             565             570             575
431 Thr Leu Lys Lys Tyr Ile Glu Gly Glu Ile Lys Asn Asn Glu Thr Asn
432         580             585             590
435 Val Phe
439 <210> SEQ ID NO: 5
440 <211> LENGTH: 1344
441 <212> TYPE: DNA
442 <213> ORGANISM: Necator americanus
444 <400> SEQUENCE: 5
445 ctcgtgccga attcggcacg agctccattc atcatgcagc gatcattcct acttctactt       60
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 11/08/2004
PATENT APPLICATION: US/10/825,692   TIME: 16:09:14

Input Set : A:\03740007aa.txt
Output Set: N:\CRF4\11082004\J825692.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:51; N Pos. 27,353,366,394,413

Invalid <213> Response:
Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:65,66,70,71,72,73,74,75,78,79,80,81

VERIFICATION SUMMARY   DATE: 11/08/2004
PATENT APPLICATION: **US/10/825,692**   TIME: 16:09:14

Input Set : **A:\03740007aa.txt**
Output Set: **N:\CRF4\11082004\J825692.raw**

L:21 M:270 C: Current Application Number differs, Replaced Current Application No
L:21 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:4143 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:0
L:4153 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:300
L:4155 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:360